ED410231 1996-11-00 Pitfalls of Data Analysis. ERIC/AE Digest.

ERIC Development Team

www.eric.ed.gov

Table of Contents

If you're viewing this document online, you can click any of the topics below to link directly to that section.

ERIC Identifier: ED410231
Publication Date: 1996-11-00
Author: Helberg, Clay
Source: ERIC Clearinghouse on Assessment and Evaluation Washington DC.

Pitfalls of Data Analysis. ERIC/AE Digest.

It can be surprising to naive observers that statistics from different research studies concerning the same issue can produce very dissimilar or contradictory results. In order to resolve this paradox, many people conclude that statistics are not actually reliable indicators of reality. Those versed in statistics, however, understand that statistics rely on assumptions. Proceeding from different assumptions, or claiming assumptions which don't apply to the research situation in question can lead to divergent results. This

digest attempts to warn against the frequent misuses and abuses of statistics. Although these issues are familiar to most statisticians, they can be easily overlooked. These problems can be considered in three broad classes of statistical pitfalls: sources of bias, errors in methodology, and misinterpretation of results.

SOURCES OF BIAS

Statistical methodology assists researchers in making inferences about a large group (a population) based on observations of a smaller subset of that group (a sample). Sources of bias are conditions or circumstances which affect the external validity of statistical results. Thus, in order for a researcher to make legitimate conclusions about the specified population, two characteristics must be present within the sample: representative sampling and valid statistical assumptions.

Representative sampling is one of the most fundamental principles of inferential statistics. This type of sampling implies that the observed group has similar characteristics to the target population in all areas that are relevant to the study. Representative sampling is necessary to make valid inferences made about the target population. Unfortunately, representative sampling can be difficult to achieve. The ideal sample is chosen by selecting members of the population at random, with each member having an equal probability of being selected for the sample. When randomization is not possible, researchers usually try to choose a sample in which their group of subjects "parallels" the population with respect to the characteristics that are thought to be important to the particular investigation.

Statistical assumptions made about various aspects of the problem determine the statistical procedure's validity. This means that certain aspects of the measured variables must conform to assumptions which underlie the statistical procedures to be applied. For example, well-known linear methods such as analysis of variance (ANOVA) depend on the assumptions of normality and independence. While assumption of normality implies that the scores in each treatment group are distributed in a way that corresponds to the so-called "normal" (or Gaussian) distribution, the assumption of independence indicates that each of the subject's scores are uninfluenced by the scores of anyone else who was tested.

ERRORS IN METHODOLOGY

There are a number of ways that statistical techniques can be misapplied to problems in the real world. These types of errors can lead to invalid or inaccurate results. Three of the most common hazards are designing experiments with insufficient statistical power, ignoring measurement error, and performing multiple comparisons.

Two types of errors can occur when making inferences based on a statistical hypothesis test: a Type I error happens if the null hypothesis is rejected when it should not be (the probability of this is called "alpha"); and a Type II error results from the failure to reject a

null hypothesis when you should (the probability of this is called "beta"). Statistical power refers to the probability of avoiding a Type II error and depends on the ability of one's statistical test to detect true differences of a particular size. The power of the test generally depends on four things: the sample size, the desired detectable effect size, the specified Type I error rate, and the variability of the sample. Based on these parameters, the power level of the experiment can be calculated. Nevertheless, the researcher can also specify the desired power level (e.g. .80), the Type I error level, and the minimum effect size which would be considered "interesting." (See Cohen, 1988, for more details on power analysis.).

If there is little statistical power, a researcher risks overlooking the effect which he/she is attempting to discover. This is especially important if one intends to make inferences based on a finding of no difference. However, it should be noted that it is possible to have too much statistical power. If the sample is too large, nearly any difference, no matter how small or meaningless from a practical standpoint, will be "statistically significant." This occurrence can be particularly problematic in applied settings, where important decisions are determined by statistical results.

Studying the relationship of multiple variables is especially troublesome because the desired knowledge is complex in nature and many different combinations of factors need to be examined. The best strategy to check these different combinations of factors is to rerun the experiment and see which comparisons show differences in both groups (also known as replication). Although this method is not irrefutable, it should provide a good notion of which effects are real and which are not. If replication is not a possibility, cross-validation--a technique which involves setting aside part of the sample as a validation sample--can also be helpful. In this system, the statistics of interest are computed on the main sample and are checked against the validation sample to verify that the effects are real. Using this technique, results that are spurious will usually be revealed by the validation sample.

Most statistical models assume error free measurement, at least of independent (predictor) variables. However, measurements are seldom perfect. Therefore, close attention must be paid to the effects of measurement errors. This is especially important when dealing with noisy data such as questionnaire responses or processes which are difficult to measure precisely.

Methods are available for taking measurement error into account in some statistical models. In particular, structural equation modeling allows one to specify relationships between "indicators," or measurement tools, and the underlying latent variables being measured, in the context of a linear path model. For more information on structural equation modeling and its uses, see Bollen (1989).

PROBLEMS WITH INTERPRETATION

In addition to difficulties with bias and methodology, there are a number of problems which can arise in the context of substantive interpretation as well. These problems usually involve determining the significance of certain findings, avoiding confusion between precision and accuracy, and unraveling the causal relationships among variables.
The difference between "significance" in the statistical sense and "significance" in the practical sense continues to elude many consumers of statistical results. Significance (in the statistical sense) is really as much a function of sample size and experimental design as it is a function of strength of relationship. With low power, a researcher may overlook a useful relationship; with excessive power, one may find microscopic effects that have no real practical value. A reasonable way to handle this sort of thing is to cast results in terms of effect sizes (see Cohen, 1994)--that way the size of the effect is presented in terms that make quantitative sense.

Precision and Accuracy are two concepts which seem to get confused frequently. It's a subtle but important distinction: precision refers to how finely an estimate is specified, whereas accuracy refers to how close an estimate is to the true value. Estimates can be precise without being accurate, a fact often glossed over when interpreting computer output containing results specified to the fourth or sixth or eighth decimal place. Therefore, one should not report any more decimal places than he/she is fairly confident of reflecting something meaningful.

Assessing causality is the reason of most statistical analysis, yet its subtleties escape many statistical consumers. For one to determine a causal inference, he/she must have random assignment. That is, the experimenter must be the one assigning values of predictor variables to cases. If the values are not assigned or manipulated, the most one can hope for is to show evidence of a relationship of some kind. Observational studies are very limited in their ability to illuminate causal relationships.

Now, of course, many of the things that are of interest to study are not subject to experimental manipulation (e.g. health problems/risk factors). In order to understand them in a causal framework, a multifaceted approach to the research (you might think of it as "conceptual triangulation"), the use of chronologically structured designs (placing variables in the roles of antecedents and consequents), and plenty of replication is required to come to any strong conclusions regarding causality.

SUMMARY

In this paper, some of the trickier aspects of applied data analysis have been discussed. In future research or data analysis, people should be certain of the following:
The sample is representative of the population of interest.

The right amount of power should be included.

The best available measurement tools should be used. If there are errors in the measures, that fact must be taken into account when interpreting the results.

Multiple comparisons need to be watched closely. If many test need to be done, replication or cross-validation should be used to verify the results.

The objective of the study should remain the focus when interpreting the data. Therefore, magnitudes rather than p-values should be studied so that one isn't seduced by "stars in the tables."

Numerical notation should be used in a rational way. This will help to avoid confusion between precision and accuracy.

The conditions for causal inference should be understood.

If causal inference must be made, random assignment should be use. In the absence of random assignment, much effort will be needed to uncovering causal relationships, requiring a variety of approaches to the question.

Although errors and misconceptions about statistical information are difficult to avoid, one can use the above suggestions to help present the information in the clearest way possible.

Mr. Helberg can be reached at SPSS, Inc., 444 N. Michigan Avenue, Chicago, IL 60611; or via e-mail at chelberg@spss.com

ADDITIONAL READING

Bollen, K. (1989). Structural Equations with Latent Variables. New York: John Wiley & Sons.
Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1990). Things I have learned (so far). American Psychologist, 45, 1304-1312.

Huff, D. (1954). How to Lie with Statistics, New York: W.W. Norton & Co.

Paulos, J.A. (1988). Innumeracy: mathematical illiteracy and its consequences. New York: Hill & Wang.

Tufte, E.R. (1983). The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press.

This publication was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract RR93002002. The opinions expressed in this report do not necessarily reflect the positions or policies of OERI or the U.S. Department of Education. Permission is granted to copy and distribute this ERIC/AE Digest.

Title: Pitfalls of Data Analysis. ERIC/AE Digest.
Document Type: Information Analyses---ERIC Information Analysis Products (IAPs) (071); Information Analyses---ERIC Digests (Selected) in Full Text (073);
Available From: ERIC Clearinghouse on Assessment and Evaluation, 210 O'Boyle Hall, The Catholic University of America, Washington, DC 20064; toll free telephone: 800-464-3742.
Descriptors: Causal Models, Comparative Analysis, Data Analysis, Error of Measurement, Research Design, Research Methodology, Research Problems, Sampling, Statistical Bias, Statistical Significance, Statistics, Test Results, Validity
Identifiers: Accuracy, ERIC Digests
###

[Return to ERIC Digest Search Page]